

# 大数据驱动下中国 CPI 统计改革的机遇和挑战

陈立双<sup>1</sup> 杨 灿<sup>2</sup>

(1. 湖北经济学院 旅游与酒店管理学院 武汉 430205;

2. 厦门大学 经济学院 福建 厦门 361005)

**摘 要:**居民消费价格指数作为衡量通胀水平的重要代表性指标,在迎来历史性改革机遇的同时也遭遇了严峻挑战。基于大数据典型特征与相关技术,系统剖析中国居民消费价格指数编制方法与公布的新问题,据此厘清其改革发展趋势、机遇与挑战,将有助于全面推进大数据时代中国居民消费价格指数编制方法与实践创新,进而促进中国居民消费价格指数统计工作的全面性、及时性和科学性,以更好地服务于大数据时代下中国的国计民生。

**关键词:**大数据; CPI 编制; 统计改革

中图分类号: F726 .C813 文献标志码: A 文章编号: 1674 - 4543(2019)04 - 0017 - 09

DOI:10.16537/j.cnki.jynufe.000427

## 一、引言

人类进入 21 世纪以来,随着互联网和信息技术的快速发展,商品交易的时间、空间和速度都实现了空前突破,同时其交易信息被系统、全面地记录下来,形成了类型繁多、时效性高的大数据。作为与数据长期打交道的政府统计部门,顺应时代发展要求,改革相关统计工作也就成为了必然趋势。目前,美国劳工统计局已开始利用网络爬虫方法采集商品特征信息,以用于居民消费价格指数(简称“CPI”) Hedonic 方法的质量调整<sup>[1]</sup>;新西兰统计局已着手利用网络抓捕数据研究价格指数的编制问题;英国和印度等国家统计局也开始研究大数据背景下 CPI 的改革与创新问题。就中国而言,习近平总书记在中共中央政治局第二次集体学习时不失时机地强调,大数据发展日新月异,应该审时度势、超前布局,推动实施国家大数据战略,加快建设数字中国,更好地服务我国经济社会发展和人民生活改善。可见,契合国家大数据发展战略和 CPI 编制的国际最新趋势,改革和创新中国 CPI 统计方法已成为事关国计民生的重大问题。

大数据时代,统计部门借助信息化手段可以弥补 CPI 统计数据缺口,扩展政府价格统计的范围,创新价格指数统计方式和提升统计测量的准确性与科学性,推进价格指数经济运行的及时性和精准性。但大数据与传统调查数据的显著差异,可能对政府现有 CPI 统计模式、流程与规范等造成许多不确定性影响,学术界却尚未对此做出相对明确的阐述和系统性回应。基于此,有必要密切联系大数据特征,契合大数据发展趋势,系统评估大数据对 CPI 编制的重要影响,以及在阐述大数据应用对 CPI 编制带来历史性机遇的同时,深入探讨可能面临的困难和挑战,并尝试提出解决的对策方案和改革措施,以促进大数据在 CPI 中的广泛应用,优化大数据时代中国 CPI 的编制效率及其数据质量。

收稿日期: 2018 - 11 - 25

基金项目: 国家社会科学基金项目“大数据背景下线上 CPI 编制理论、方法与应用研究”(16BTJ028)

作者简介: 陈立双(1976—),男,湖北监利人,湖北经济学院旅游与酒店管理学院副教授,博士,主要研究方向为经济统计和旅游统计;杨灿(1957—),男,湖南长沙人,厦门大学经济学院教授,博士,博士生导师,主要研究方向为统计理论方法、国民经济核算和宏观经济分析。

## 二、大数据时代中国 CPI 统计方法改革的若干机遇

## (一) 大数据记录技术应用所带来的机遇

## 1. 创新中国 CPI 基础性数据采集方法 破解传统数据采集瓶颈

基础性数据采集是整个 CPI 统计的源头,是确保 CPI 数据质量的关键性环节,在整个 CPI 统计过程中举足轻重。尽管国家统计局长期坚持数据调查方法的改革与创新,但根据目前的现实情况来看,中国 CPI 基础性数据的调查形势依然严峻,具体表现为:(1)对数据采集长期大量的时间和精力投入,严重制约了数据采集效率;(2)基层统计调查队承担着繁杂的统计调查任务,统计队伍及其专业素质面临严峻考验;(3)商品权重数据的家庭入户调查存在配合程度低,以及瞒报、漏报等突出问题,而且相应支出数据与规格品价格数据还存在匹配性问题等。由此可见,中国 CPI 传统数据采集效率相对较低,采集数据质量也不容乐观。

大数据技术可以大大缩短 CPI 数据采集时间,提高数据采集质量。实践中,统计部门可以利用居民网络消费大数据的采集为突破口,借助大数据记录技术,利用传感器、电子商务、移动互联网、穿戴工具等新技术和手段,将 CPI 基础性数据的采集以“人工采集”为主逐步过渡到以“信息化采集”手段为主,从而实现 CPI 基础性数据采集方法的改革。表 1 列出了中国 CPI 可能借助大数据记录技术改进数据采集方法的相关应用。

表 1 CPI 传统数据采集与大数据采集技术应用比较

	传统数据	大数据
价格数据获取	商场、超市、农贸市场、服务网点等抽样调查规格品价格	基于移动互联网和各种信息技术设备等实时记录线上各交易平台与线下各交易网点消费品交易的各种信息
权重数据获取	居民家庭消费支出的抽样调查	借助各种穿戴设备、网络支付和银行卡支付等方式,采用专业化的信息技术采集手段实时、详细地记录住户各消费支出信息

借助大数据记录技术改进数据采集的重要意义在于,一方面可以实时同步获取居民消费品价格数据与权重数据,以解决传统 CPI 中商品价格与权重数据不匹配的矛盾;另一方面,利用信息技术手段代替人的因素,以逐步实现对 CPI 数据采集工作的智能化,并在大幅降低其数据调查成本的同时,有效缓解传统 CPI 数据采集水平相对滞后,以实现居民消费数据采集效率和质量的提升。

## 2. 契合信息技术 扩展居民消费数据采集范围

大数据技术扩展居民消费数据采集范围主要体现在以下层面:(1)对于线下交易。首先,运用大数据和云技术构建政府大数据共享平台<sup>①</sup>,并通过共享工商管理、税务部门等相关企业的企业登记信息,以全面掌握商品零售网点;其次,运用大数据技术手段将原有抽样调查方法分类,稳步有序地发展为依靠信息化手段为主,并最终实现对地区所有商场、医院、超市、农贸市场、服务网点等交易信息的全面调查<sup>②</sup>;(2)对于线上交易,实现网络全覆盖式数据的全面采集。初步考虑,在各网络交易和电商平台等,运用网络爬虫和抓捕技术全面采集 CPI 统计所需要的相关数据。对于这一方法,美国劳工统计局和新西兰统计局已着手利用网络抓捕数据研究价格指数的统计问题。进一步考虑,通过大数据技术和云平台,深入推进统计部门与线上交易平台、电商等相关企业部门开展网络交易大数据的共享等<sup>③</sup>。总之,大数据调查技术在 CPI 数据采集中广泛、系统应用,必将极大地扩展居民消费数据的统计

①2015年8月31日,国务院印发的《促进大数据发展行动纲要》就提到,2018年底前,建成国家政府数据统一开放平台。2020年底前,逐步实现政府各部门、各事业单位等民生保障服务相关领域的政府数据集向社会公众开放。

②目前,不少欧盟国家已经采用立法和信息技术手段等多重方式,实现统计部门对超市扫描数据的共享。

③目前这一举措的实施难度较大,而且还需获得立法支持。但居民网络消费大数据是事关国计民生的重要战略性资源,在相关配套措施得以保障的情况下,政府与业界间相关数据的共享将成为一种必然趋势。

来源和范围,显著改善 CPI 基础性数据的代表性及其质量。

### 3. 发挥大数据便捷性优势,破解服务消费品数据采集难题

《消费者价格指数手册:理论与实践》(以下简称《CPI手册》)将服务消费品的采价问题列为六大值得关注的问题之一,并指出尽管服务业已变得极为重要,但在消费者价格指数中仍未获得足够重视<sup>[2]</sup>。对中国而言,CPI中服务消费品的采价也是一大棘手的历史性难题。随着我国第三产业的快速发展和信息技术的广泛应用,居民在旅游、教育、娱乐、医疗保健、养老、美容等方面的服务消费支出均呈现爆发式增长,但目前服务业统计水平依然相对滞后,服务消费类商品在CPI中的权重问题日益凸显。究其原因主要有:(1)服务业覆盖面广,布局分散,构成复杂,规模参差不齐,部分服务业还具有隐蔽性、不规范性甚至非法性等<sup>[3]</sup>;(2)服务行业发展快,服务企业名录库更新相对滞后;(3)部分调查对象账目不健全、抽样调查配合程度差等。

针对上述问题,可以基于即将建成的“国家政府数据统一开放平台”系统、实时更新服务类企业名录库,在进行行业细分类的基础上,针对服务网点和相关服务企业,通过查看现场视频扫描、查询平台支付信息、连续跟踪等方式,有针对性地采集服务类消费品交易数据;其次,通过大数据分析审核技术,针对重点行业和企业开展统计数据质量监督和评估工作,以逐步优化和完善线上、线下服务类消费品交易价格与支出数据的信息化采集。实际上,早在2014年,挪威在进行深入研究的基础上,就将相关在线服务消费品分为旅游或假期住宿(Travel or Holiday Accomodation)与其它在线服务(Other Services)两大类,并尝试集中采集在线服务消费品大数据。

## (二) 大数据及其关联技术优化价格指数方法的机遇

### 1. 运用基于大数据下的机器学习进行代表性规格品的有效甄别

规格品的代表性和时效性对CPI数据质量有着举足轻重的影响。目前,中国传统CPI规格品的确定还存在一些弊端,如服装类规格品的确定,存在多变性与调查方式之间的矛盾<sup>[4]</sup>;又如,当前电子技术高速发展,电器类产品升级换代周期大大缩短,经常会出现规格品代表性失效或已被市场淘汰等情况<sup>[5]</sup>。借助大数据强大的计算技术优势,可以对现有CPI规格品的选取方法进行改革。例如,可借助大数据机器学习这一先进技术,对全部消费类商品进行智能化分类的基础上,精准把握各类商品生存周期、新旧产品变更频率、商品交易量变化规律与趋势,再按照一定程序,在尽可能满足可比性和稳定性的基础上,借助无监督学习最终识别出更具代表性的规格品。表2列出了中国传统CPI规格品选择方法及其可能的改革措施。

表2 中国传统CPI规格品确定方法及其大数据计算技术应用下的改革与影响

数据来源	规格品数量	规格品甄选方法与标准	改革前后相关对比
传统抽样调查数据源	600种以上	消费量大、价格变动代表性强、性质差异大、生产销售前景好、合格品等标准	规格品更新频率低,代表性偏差难以准确把握
大数据技术采集数据源	以机器学习确定数目为主	尽可能满足规格品可比、稳定性的条件下,采用机器学习(如借助无监督学习方法,K-means clustering等)确定规格品	有效轮换规格品,把握规格品的代表性偏差,提升其代表性

在大数据驱动下,应用机器学习甄别规格品,可以简化分类程序、提供更为方便的测试、验证和系统维护潜力。上述规格品确定方式的重要转换,将极大地优化规格品的代表性和可比性。

### 2. 利用大数据可视化技术优化CPI指数发布

尽管目前国家统计局关于CPI数据的对外发布,内容丰富,数据相对真实可靠,透明度相对较高,但在大数据时代,其发布方式也存在一些突出问题。例如,数据的发布倾向于以简单展示为主,界面形式相对固定,对于数据间的交互式特征、关联性、趋同性、动态比较等缺乏适当的体现,数据用户与统计部门之间缺乏必要的信息交流和反馈渠道。

大数据可视化技术是关于数据视觉表现形式的一种科学手段,借助大数据可视化技术将有助于提高指数发布的技术和方法,化解传统CPI数据对外发布存在的不足。首先,它可以通过可视化框架和灵活的数据转换,满足不断变化的信息需求,促使数据的展示更具可行性、可塑性和智能性。其次,可视化技术能够在展示价格指数信息的同时,根据使用者需求调用相关数据,并经过简单的指令化处理后输出目标指数,满足不同用户的多层次、多目标需求。最后,可视化技术系统能够及时收集和存储客户反馈的有关信息,以便统计部门了解数据用户对价格指数的应用需求与建议,从而进一步完善和优化指数体系及其数据服务功能。

### (三) 大数据海量信息优化指数编制方法的机遇

#### 1. 借助大数据海量信息全面优化中国CPI权重问题

《CPI手册》列出了标准指数方法中六大值得关注的突出问题,其中就有三大问题与权重问题密切相关。就中国CPI而言,实际上还存在其它权重问题。例如,CPI汇总的权重数据与商品价格数据来源渠道不一致,各小类商品权重数据尚未对外公布,以及由此引发的误解和质疑等。大数据时代,系统同步记录的商品交易价格与数量大数据,将有助于解决CPI编制中的各类权重问题,这具体表现在:(1)低层汇总可采用加权价格指数,低层汇总替代偏误将会随之大幅下降;(2)中国权重基期每五年调整一次,据此使用的价格指数事实上并非拉氏指数,而是Young指数。而大数据驱动下权重数据的实时获取,有望化解支出权重与价格参考时间不一致,即指数的权偏误问题;(3)有望对外系统发布代表性规格品及各级各类指数的权重数据。总之,利用大数据海量信息化解上述系列权重问题,不仅可以有效解决指数替代和权偏误,而且还可以回应民众相关权重问题的质疑,提升CPI数据质量和政府统计部门公信力。

#### 2. 运用大数据海量信息优化季节性产品处理和商品质量调整方法

在现有CPI基础数据采集方式下,国家统计局尚未对季节性产品和商品质量信息进行相对全面的采集,导致CPI编制中季节性产品处理方法选择上的困难和应用效果评估上的难以开展、商品质量调整工作推进缓慢。事实上,这不仅是中国CPI编制亟需解决的重要实践任务,也是国际上悬而未决的棘手难题。借助大数据技术采集海量信息,尤其是各种商品交易量及其相关商品特征的详细信息,将迎来推进季节性产品处理和商品质量调整工作的历史性机遇。相应地,利用大数据可深入推进的相关实践统计工作主要包括:(1)厘清各类季节性产品在相应类别商品中的分布情况,详细了解季节性商品所占比重;(2)系统考察季节性产品的变动趋势和周期性的变化规律;(3)深入探讨季节性产品对价格指数波动的重要影响,比较各种季节性产品处理方法的应用效果,寻求中国CPI更为合理的季节性产品处理方法;(4)全面推进商品特征信息的采集及其数据库的构建工作、系统厘清各类商品质量变化特征;(5)结合商品质量变化特征,基于商品质量调整的各种方法,适时开展商品质量显性和隐形调整的理论研究与实践统计工作;(6)深入推进Hedonic方法在中国CPI中的重要实践应用等。

不难发现,CPI中大数据海量信息的有效开发与应用,将从根本上化解中国CPI编制中季节性产品问题和商品质量变化问题。

#### 3. 利用大数据海量信息编制不变价增加值核算的相关指数

目前中国多数服务行业不变价增加值核算使用的是缩减法,其结果准确性依赖于缩减指数的数据质量及其是否合理应用。实践中,尽管中国服务行业分类较为详细,但大多缩减指数依然统一采用CPI或CPI中的分项价格指数加以代替,其代表性和针对性明显不够<sup>[6]</sup>。利用大数据的海量信息,同时借助政府行政管理中记录的大数据信息,编制细分类目的服务品价格指数,将有助于提高缩减指数的代表性、匹配性和准确性,进而促进服务业不变价增加值数据质量的提升。表3列出了中国部分服务业不变价增加值核算中,相关指数应用存在的问题,以及可补充编制的细分消费类商品价格指数。

表 3 部分服务行业缩减指数问题与细分类指数的补充编制

不变价增加值类目	所使用的替代指数	指数存在问题	补充编制的指数
餐饮业	商品零售价格指数	缺乏细分类服务指数	在外用膳商品价格指数或外卖价格指数
物业管理、房地产中介和其他房地产活动	居民自有住房用固定资产投资价格指数替代; 其他用 CPI 服务价格指数替代	自有住房、物业和中介服务缩减指数不匹配	自有住房、物业和中介等服务类价格指数
金融业	金融业缩减指数	银行业产出与该指数匹配性差	银行业其他服务类项目消费指数
公共管理	CPI 指数	指数的应用粗糙	居民公共服务消费价格指数
文化、体育和娱乐业	娱乐教育文化用品及服务品价格指数	市场性项目的指数匹配性差	居民在线文化、体育和娱乐消费细分类价格指数

#### 4. 利用大数据海量信息创新中国 CPI 理论与方法的经济计量研究

指数理论与实践问题经历了 300 多年的研究和探索,形成了丰富系统的指数基本理论,并积累了大量实践经验,但 CPI 编制实践中依然存在不少尚待解决的难题。例如: CPI 各类偏误、链式指数的漂移性、商品间相互替代弹性及其时变情况、消费者效用水平的动态变化特征、消费者偏好时变情况、商品价格粘性规律等。对于这些问题,国外学术界借助超市扫描数据展开了相对深入的经济计量研究,但由于居民消费习惯、收入水平和市场特征等因素的影响,导致上述问题在各国呈现出不同的表现形式和变化规律。就国内学术界而言,关于上述问题的研究尚未引起足够重视,其相关问题的计量研究也较为滞后。借助大数据海量信息,围绕上述系列中国 CPI 编制的现实问题,创新性地开展指数理论与方法的经济计量研究,无疑将为中国 CPI 编制提供强大的理论与方法支撑。

#### 5. 运用大数据海量信息全面创新特型 CPI 指数编制工作

总体来看,目前国家统计局编制的 CPI 主要基于省(区)、市(县)等不同行政区划单位的 CPI,再汇总得出全国 CPI。这一 CPI 体系主要反映了相应地区和行政区划范围内的价格变化水平,其平均的方式以地域性特征为主,其统计标志显得较为单一。大数据背景下,统计部门可以参考更多标志、满足多方需求,编制多层次、多视角地反映商品价格变化水平和居民消费偏好的特型价格指数,从而实现价格指数功能的进一步扩展以及指数代表性和精准性的实质性提升。基于此,表 4 列出了部分标志下特型 CPI 编制的初步设想。

表 4 不同标志下特型价格指数编制的设想

消费品属性	消费者特征	地域发展差异	支付方式和场景
核心 CPI	不同收入阶层居民 CPI	东部地区 CPI	网上银行
菜篮子 CPI	不同年龄阶段居民 CPI	中部地区 CPI	手机银行
耐用消费品 CPI	不同文化程度居民 CPI	西部地区 CPI	微信银行
非耐用消费品 CPI	不同家庭构成 CPI	少数民族地区 CPI	预付卡
季节性产品 CPI	婴儿消费品 CPI	一线城市 CPI	POS 机支付
奢侈品、必需品 CPI	老人及其保健品 CPI	二、三线城市 CPI	第三方支付

借助大数据海量信息创新特型 CPI 指数体系,既是进一步完善 CPI 在反映商品价格变化差异、居民消费模式与偏好差异、地区经济发展差异、商品价格销售渠道和方式差异等功能方面的重要举措,也是 CPI 契合大数据和信息经济发展的必然要求。特型 CPI 指数的全面创新,将进一步优化 CPI 的代表性、完善 CPI 体系构架、促进 CPI 宏观经济功能向微观领域的扩展与深化。

#### (四) 运用大数据的实时性提高指数的时效性

##### 1. 运用大数据的实时性优化规格品及其权重,以确保商品篮子的时效性

中国 CPI 调查基本分类在一个基期年(5 年)内保持不变,基期年内权重数据每年仅微调。首先,

这样的更新频率与国际(上尤其是部分欧盟国家)还存在较大差距;其次,大数据时代,随着科技快速发展和企业间竞争日趋激烈,产品更新换代越来越快<sup>[7]</sup>,销售前景好的产品几乎每天都会发生变化<sup>[8]</sup>。因此,忽略规格品及其权重更新问题将会造成CPI的严重偏误<sup>[9]</sup>。尤其是,随着近些年中国居民收入水平的不断提高,老百姓消费结构调整步伐日益提速,这一问题也越来越突出。

大数据背景下,居民购买行为被系统记录下来,并产生了实时更新的大数据,新旧产品变更情况及各类商品权重变化情况也被及时捕捉。借助大数据的这种实时性,不仅可以对旧消费品的退市和新消费品的入市情况进行动态监测,进而选择更具时效性的代表性规格品,还可以优化商品各层次的分类,实现代表性规格品权重的实时更新,以更好地反映居民消费模式和偏好。可见,运用大数据的实时性,为进一步优化商品篮子的时效性和代表性提供了重要历史性机遇。

## 2. 运用大数据的实时性提高统计指数及经济运行信息的及时性

统计的及时性是指统计数据从调查到发布的时间间隔,是评估统计工作质量评价标准,其时间间隔越短,及时性也越强<sup>[10]</sup>。而经济运行信息的及时性是指对已经发生的经济运行情况或者重要经济活动进行确认、计量和报告的时间间隔。

首先,关于CPI统计数据的及时性。大数据的快速反应是其重要价值所在<sup>[11]</sup>,大数据能够实时记录居民消费信息,瞬时完成数据采集工作,并利用云技术的快速处理编制更高频率的价格指数,从而大大缩短CPI从数据调查到发布的时间间隔。其次,关于CPI对经济运行信息的及时性。大数据背景下CPI的及时性是经济运行信息及时性得以实现的前提条件,借助更高频率的CPI指数,能够更为灵敏地确认、计量和报告经济运行信息,尤其是通胀水平和经济拐点的重要信息。事实上,阿里研究中心编制的网络零售价格指数<sup>[12]</sup>、McLaren N等<sup>[13]</sup>及张崇等<sup>[14]</sup>的研究都表明,基于大数据编制的价格指数较传统CPI具有时间上的领先优势,能够更为及时地预测宏观经济拐点。

## 三、大数据背景下中国CPI编制面临的挑战

居民消费大数据作为一种全新的数据资源,在为创新中国CPI统计工作提供重要历史性机遇的同时,也由于大数据的多源性、涌现性、巨量性等特征,对统计部门利用大数据资源编制CPI带来了一些挑战。

### (一) 大数据资源整合与共享的挑战

随着信息技术的快速发展和物联网的日益普及,商业模式日趋多元化,各种各样的商业记录数据、互联网爬虫数据和超市扫描数据、零售网店穿戴工具支付的数据呈现爆发式增长。这些种类繁多、来源不同的居民消费大数据分散在各个不同的区域和平台,形成了所谓的“信息孤岛”。为提高大数据应用价值,就需要系统整合这些多源头的居民消费大数据,进而实现消费大数据资源的有效共享。但由于涉及到商业机密、数据隐私泄露风险、竞争与合作博弈等诸多深层次问题,使得业界与政府在居民消费大数据方面的合作与共享进展缓慢;另一方面,由于作为网络空间中唯一客观存在的大数据有多种格式<sup>[15]</sup>,这些格式可能与政府CPI基础性数据在统计标准、统计口径和统计规范上存在较大差异,甚至冲突。这些都将为居民消费大数据资源的整合和价值的深入挖掘带来困难。因此,大数据资源整合与共享的挑战,是大数据背景下CPI编制需要解决的首要问题。

针对该问题,政府可以通过顶层设计逐步推动相关企业和部门大数据平台与官方大数据平台实现某种形式的对接,并据此设计大数据资源整合管理体制,针对不同源头消费大数据制定资源整合与技术标准,构建统一的居民消费大数据采集、存储和挖掘平台。

### (二) 随机性与涌现性对政府CPI统计生产方式的挑战

传统CPI基础性数据来源于抽样调查获得的小样本数据,其数据按照事先设计的统计模式和制度规范搜集,数据具有体量稳定、连续性强和高度可控性等显著优点。而居民消费大数据则是智能芯片、传感器等自动采集或者消费者无意留下的,相对于传统CPI基础性数据而言,其生成具有明显的随机性、不稳定性、次生性以及复杂性、不确定性和涌现性<sup>[16]</sup>,从而对政府CPI统计生产方式提出了一

定挑战:(1)对政府统计数据采集方式的挑战。大数据的生成与记录都离不开互联网和信息技术的强有力支持,这就决定了大数据的采集需要以互联网和信息技术为媒介,而传统的随机抽样方法将不再有效。(2)对数据存储工作的挑战。大数据的爆发式涌现决定传统的纸质和硬盘等媒介难以完成其存储工作,需要借助云端和大数据平台才能够完成其存储任务。(3)对数据统计标准的挑战。对于统计部门而言,居民消费大数据不是基于 CPI 编制目的生成的,而是属于次生数据,这些数据与传统抽样调查获取的数据在统计内容与范围、统计边界与频率、统计规范性上都存在显著差异,从而对 CPI 统计标准带来了挑战。(4)对统计方法的挑战。大数据爆发式涌现伴随着太多噪声和大量冗余信息,从而使其价值密度大大降低,这对传统调查下的“小样本数据”统计方法来说是一个不小的挑战。对此,需要借助云计算技术,实现大数据的充分挖掘和有效萃取,以实现将传统的经验归纳法演绎为“大数据归纳法”,达到“以大见小”的统计目的。

### (三) 实时更新的高频率大数据对价格指数构造方法的挑战

居民消费大数据实时更新,属于典型的高频数据,为了尽可能提高商品的匹配性,高频率的链接是必不可少的<sup>[17]</sup>。而困扰学术界的一大难题是,高频率数据下商品销售价格与交易数量的大幅波动,导致了传统链式价格指数出现链式漂移<sup>①</sup>,包括日频 Törnqvist 链式价格指数<sup>②</sup>、周频 Törnqvist 链式指数以及月频 Fisher 链式指数等都产生了一定程度的链式漂移,而且其漂移性扭曲程度还随时间的累积日趋严重。因此,大数据背景下高频指数的编制,对传统价格指数构造方法带来了严峻挑战。针对这一问题,Avancic L 等首次将具有循环性特征的国际 GEKS 多边比较指数引入 CPI 编制方法中,并将其研究成果于 2011 年正式发表在《Journal of Econometrics》上<sup>[18]</sup>,该创新研究在国际学术界获得了高度认可。但根据既有研究来看,其构造的滚动 GEKS 指数并不理想,还存在指数序列扩展方法选择、窗口长度的选择、恒等性检验所造成的偏误、难于向公众解释等问题。大数据背景下,高频数据价格指数构造问题至今悬而未决,其有效解决还有待于理论与实务界的继续深入研究。

## 四、大数据背景下中国 CPI 编制的重要启示

尽管大数据与 CPI 指数编制方法的结合已成为一种必然趋势,但这一创新统计工作显然比传统 CPI 编制更为复杂,技术要求也更高,同时其相对规范、科学、有效的编制程序的建立也是一个不断摸索和逐步完善的过程。在这一过程中,统计部门需要积极开展相关评估、实践尝试和研究工作,以充分挖掘和有效扩展大数据背景下 CPI 的创新功能与重要经济价值。

### (一) 正确评估大数据对中国 CPI 编制工作的重要影响

大数据背景下,居民消费数据来源发生了质的变化,并对中国传统 CPI 统计方法产生了重大影响。具体表现为:(1)近些年线上消费呈现快速发展之势,居民消费方式在信息技术的支持下也发生了巨大变化,信息化的支付手段与传统相对落后的数据采集方式形成了鲜明的对比;(2)电商企业、咨询机构、科研院所等发布的价格指数在社会上已经形成了一定的影响力,并对统计部门 CPI 数据的权威性和公信力带来了一定程度的挑战。只有契合居民消费环境和消费行为已经发生根本性转变这一趋势,并有效评估大数据对中国 CPI 编制工作可能带来的重要影响,才能有效把握大数据背景下中国 CPI 编制的重要历史机遇,抢占大数据时代 CPI 统计的国际制高点,创新发展 CPI 的重要经济功能。

### (二) 与业界开展合作推进中国 CPI 编制向大数据技术方法过渡

目前,业界在大数据云平台的搭建、大数据处理技术与统计方法等方面具有“先发”优势,能够为政府统计部门提供技术指导,而政府统计部门在指数统计方法及其实际应用等方面更具经验,能够为业界价格指数实践提供编制供专业性的指导。基于此,大数据背景下中国 CPI 指数的编制,可通过构建

① 环比指数通过连乘积而形成的链式指数与相应定基指数值结果不一致的现象,称为链式漂移。

② Watanabe K, Watanabe T(2014) 利用扫描数据编制了  $\Delta t$  为 1~365 天的不同频率的链式指数,结果发现, $\Delta t$  越大,链式漂移性扭曲程度越严重;以天,链式指数漂移性问题都相当严重,链式指数漂移性问题明显减少。由此可见,通常情况下,传统指数构造下的月频、周频、日频链式指数都会存在较为严重的漂移性问题。

以国家统计局为主导的政企合作方式,具体包括:(1)在企业大数据技术和信息化手段的帮助下,选取个别市县的农贸市场、超市和医院等既定样本单位和网店,运用大数据技术和先进电子设备展开居民线下消费信息的实时记录与存储,待技术和条件成熟后再扩大试点范围,以实现大数据在价格调查应用方面取得突破;(2)从国家层面和战略角度继续深入推进国家统计局与业界合作编制相关价格指数,在合作的过程中系统地探讨和确立大数据背景下价格指数编制的制度规范、相关标准、统计方法和统计流程等。

### (三) 积极开展大数据技术与CPI编制相结合的基本理论与方法研究

大数据科学与技术作为一个新兴交叉性学科,在CPI编制中演绎出的理论与方法来自多个学科领域,包括统计学、数学、经济学、计算机科学与技术。深入推动大数据技术与方法在中国CPI编制中的实践应用,挖掘CPI的潜在价值,离不开多学科交叉下的基本理论与方法研究。

结合居民消费大数据基础研究,需要开展的相关理论与方法研究主要包括:(1)消费大数据的内在机理研究,包括消费大数据的结构研究及相应CPI的经济与社会价值评估;(2)居民消费大数据计算方法研究,具体包括消费大数据的复杂性、基本表示方法及CPI相关计算模型的研究;(3)居民消费大数据应用研究,即消费大数据CPI潜在价值挖掘,宏观、中观及微观经济领域的价值与功能,以及居民消费大数据的共享、安全与隐私等的研究。

## 五、结束语

正如计算机技术的普及将算盘送进博物馆、数码相机的出现加速了传统胶卷业的凋零一样,大数据作为一种新生事物同样会引发传统观念和技术的革命。大数据在居民消费领域日益广泛的应用,必定会促使传统CPI统计生产方式与流程发生重大变革。然而,借助大数据资源编制CPI,实施难度大、影响因素多、耗费时间长且牵涉面广,是当前社会各界共同面临的重大课题。

目前国际上尚无成熟的经验和方法可供借鉴,相应改革也不可能一蹴而就,这就需要我们敢于立足现实,勇于突破传统模式和思维,积极适应大数据环境,努力探索大数据与CPI编制相结合的关键理论、方法及实践问题,有效扩展大数据时代中国CPI统计分析的广度和深度,促进中国CPI的准确性、及时性和科学性,以更好地服务于大数据背景下中国的国计民生。

## 参考文献:

- [1] Groshen E. Opportunities and Challenges Facing the Bureau of Labor Statistics [J]. Business Economics, 2015, 50(2): 91-95.
- [2] ILO, IMF, OECD, et al. Consumer Price Index Manual: Theory and Practice [M]. Geneva: International Labour Office, 2004: xxi-xxiii.
- [3] 杨灿. 关于服务业统计若干问题的探讨[J]. 统计研究, 2009, (1): 40-44.
- [4] 刘洪仁. CPI调查服装类价格采集的困境与思考[J]. 中国统计, 2011, (10): 32-33.
- [5] 王有杰, 孙晓玲. 如何提高电器类商品价格采集质量[J]. 统计与经济, 2016, (5): 60-61.
- [6] 朱之鑫, 许宪春, 等. 中国不变价国内生产总值核算方法研究[M]. 北京: 北京大学出版社, 2012: 104-110.
- [7] 江梦琴. 我国电子商务产业基于竞争力提升的品牌塑造战略研究[D]. 广州: 广东外语外贸大学, 2015.
- [8] Watanabe K, Watanabe T. Estimating Daily Inflation Using Scanner Data: A Progress Report [R]. CARF Working Paper Series, 2014, (No. F): 342.
- [9] Melser D. Accounting for the Effects of New and Disappearing Goods Using Scanner Data [J]. Review of Income and Wealth, 2006, 52(4): 547-568.
- [10] 马建堂. 大数据在政府统计中的探索与应用[M]. 北京: 中国统计出版社, 2013: 112-118.
- [11] 徐继华, 冯启娜, 陈贞汝. 智慧政府: 大数据治国时代的来临[M]. 北京: 中信出版社, 2014: 48-50.
- [12] 阿里研究中心. 网络零售价格指数(ISPI)报告[R/OL]. 北京: 阿里研究中心, (2011-10) [2018-10



- 26 ]. <http://www.aliresearch.com/blog/article/detail/id/13316.html>.
- [13] McLaren N ,Shanbhogue R. Using Internet Search Data as Economic Indicators [R/OL]. Bank of England Quarterly Bulletin , June 2011. [2018 - 10 - 28 ]. <http://www.bankofengland.co.uk/publications/quarterly-bulletin/qb110206.pdf>.
- [14] 张崇,吕本富,彭赓,等. 网络搜索数据与 CPI 的相关性研究[J]. 管理科学学报 2012 ,15( 7 ): 50 - 59.
- [15] 李国杰. 大数据研究的科学价值[J]. 中国计算机学会通讯 2012 ,8( 9 ): 8 - 15.
- [16] 李金昌. 从政治算术到大数据分析[J]. 统计研究 2014 ,( 11 ): 3 - 14.
- [17] Haan J D , Van Der Grient H A. Eliminating Chain Drift in Price Indexes based on Scanner Data [J]. Journal of Econometrics ,2011 ,161( 1 ): 36 - 46.
- [18] Ivancic L ,Diewert W E ,Fox K J. Scanner Data ,Time Aggregation and the Construction of Price Indexes [J]. Journal of Econometrics ,2011 ,161( 1 ): 24 - 35.

责任编辑、校对: 王 旭

## The Opportunities and Challenges of China's CPI Statistical Reform under the Driven of Big Data

CHEN Li - shuang<sup>1</sup> , YANG Can<sup>2</sup>

( 1. School of Tourism and Hospitality Management , Hubei University of Economics , Wuhan 430205 , China;

2. School of Economics , Xiamen University , Xiamen 361005 , China)

**Abstract:** As a representative indicator of the inflation level , consumer price index ( CPI) has encountered severe challenges while obtaining historic opportunities of reform. Based on the typical characteristics of big data and the related technologies , the paper systematically analyzes the compilation method of China's CPI and the series of new issues highlighted by the announcement , and clarifies the important trends , new opportunities and challenges of the reform and innovation. This will help to promote the important innovations in the compilation method and practices of China's CPI in the era of big data , and further promote the comprehensiveness , timeliness and scientificity of the statistical work of CPI in China so as to better serve China's national economy and people's livelihood in the era of big data.

**Key words:** Era of Big Data; CPI Compilation; Statistical Reform